

# A Trained Encoder for Wearable Motion Sensing, Validated Honestly

The first real, trained checkpoint for VAAS-X's OIE (Operational IMU Embedding) encoder -- previously an architecture-only scaffold -- validated on human subjects held out of training, deployed to production, and streamed end to end against the live Bootstrap API.

**19**

ACTIVITIES TESTED,  
EVERY ONE SEPARATED  
ABOVE CHANCE

**23.5**

WELCH'S T-STATISTIC  
ON HELD-OUT  
SUBJECTS

**251**

REAL PRODUCTION  
STREAM REQUESTS,  
ZERO ERRORS

**0**

GPU REQUIRED

## The problem

A raw IMU stream -- accelerometer plus gyroscope, sometimes magnetometer -- is six or more correlated numeric channels with no obvious similarity metric. "Is this movement like that movement?" isn't answerable with a single channel's threshold the way one sensor reading sometimes is. The OIE encoder turns a window of raw IMU samples into a fixed-length embedding where cosine similarity is a meaningful answer to that question -- the same operational-memory pattern VAAS-X already applies to industrial telemetry, applied to motion.

## What's trained

**Architecture:** a multi-scale temporal CNN with cross-channel attention, taking  $[B, 6, T]$  windows (3-axis accelerometer + 3-axis gyroscope) and producing a 256-dimensional, L2-normalized embedding. **Training method:** SimCLR-style contrastive self-supervised learning (NT-Xent loss) -- no activity labels used during training itself. **Data:** the UCI "Daily and Sports Activities" dataset (Barshan & Altun, 2010) -- 8 real subjects, 5 Xsens IMU units per subject, 19 activities, 60 five-second segments per activity per subject at 25Hz. This checkpoint uses the torso unit's 6 accelerometer/gyroscope channels only.

**Split:** by subject, not window. Subjects p1-p6 trained (6,840 windows); subjects p7 and p8 held out entirely -- never seen during training.

## The honest validation result

Held-out windows from subjects p7 and p8 were compared pairwise: same-activity cosine similarity (mean 0.207) vs. different-activity (mean 0.017) -- separation 0.190, Welch's  $t = 23.53$ , a real effect, not noise. Every one of the 19 activities showed positive separation, but unevenly -- the strongest separation clustered around gross posture changes and high-impact movement (lying positions, jumping, running: gaps of 0.37-0.80); the weakest clustered around activities whose discriminating motion lives in the limbs, not the torso (vertical cycling, flat-treadmill walking, the stepper: gaps of 0.01-0.04). This is a real, physically explainable limitation of a torso-only sensor, reported honestly rather than averaged away.

| Activity                                                       | Same-act. sim | Diff-act. sim | Gap   |
|----------------------------------------------------------------|---------------|---------------|-------|
| Lying on right side                                            | 0.848         | 0.046         | 0.801 |
| Lying on back                                                  | 0.577         | 0.001         | 0.576 |
| Jumping                                                        | 0.387         | 0.006         | 0.381 |
| Running, treadmill                                             | 0.390         | 0.018         | 0.372 |
| ... (19 activities total, full table in the technical writeup) |               |               |       |
| Standing                                                       | 0.092         | 0.053         | 0.039 |
| Walking, treadmill flat                                        | 0.048         | 0.023         | 0.025 |

|                        |       |       |       |
|------------------------|-------|-------|-------|
| Cycling, vertical bike | 0.047 | 0.033 | 0.013 |
|------------------------|-------|-------|-------|

## Real production deployment

---

The checkpoint was deployed to the live `vaasx-bootstrap` service via a chunked transfer, with the reassembled file's SHA-256 verified byte-for-byte identical to the local checkpoint before being moved into place. A real end-to-end test followed: a held-out sample was streamed through the production streaming adapter (sliding window, matching the 125-sample/25Hz training window) against the live Bootstrap API on a throwaway device. **251 successful stream requests** (375 rows – 125-sample window + 1, exactly matching the sliding-window arithmetic), zero errors.

## Scope, honestly

---

**What's proven:** a torso-mounted 6-axis IMU stream, sampled at 25Hz in 5-second windows, produces embeddings whose cosine similarity meaningfully separates different human activities for subjects never seen during training -- validated per-activity, not just on average. The full deployment and streaming path into Bootstrap is real and exercised end to end on production infrastructure.

**What this does not yet show:** other sensor placements (the dataset has 4 more body units never trained on), the unused magnetometer axes, any sample rate or window length other than 25Hz/125 samples, any domain beyond lab-collected daily/sports activities, or anomaly detection built on top of the embedding -- that is a separate validation this encoder has not yet been through.

## Deployment model

---

Identical to every other vertical: ingestion and retrieval run against the hosted resolver behind a single API key, with no local GPU or vector database to operate. The encoder itself runs client-side, wherever the device or server is -- production only ever stores and indexes the embedding it's sent.

---

**Get started:** Free tier -- instant API key, no credit card -- `pip install vaas-x`, connect a source, query your first episode in under ten minutes.

Full guide: [vaasx.com/docs/getting-started.html](https://vaasx.com/docs/getting-started.html) | Enterprise: [hello@vaasx.com](mailto:hello@vaasx.com)